THE ROLE OF ATTENTION IN AUDIO-VISUAL INTEGRATION

Michael Schutz[1], Laura Silverman[2]
[1] School of the Arts, McMaster University (Hamilton, Ontario) schutz@mcmaster.ca
[2] University of Rochester Medical Center (Rochester, New York) laura_silverman@urmc.rochester.edu

1. INTRODUCTION

When listening to a person speak, our perceptual system effortlessly integrates the sounds detected by our ears with the gestures and lip movements seen by our eyes, giving rise to the experience of a unified event. However, this *illusion of unity* is in fact the end-result of a complex process integrating disparate information. The dominant theory for explaining this process is known as optimal integration – predicting that sensory information is weighted in proportion to its quality (Ernst & Banks, 2004). Given the superior temporal acuity of the auditory system, audition generally dominates when resolving conflicts in timing - such as judging event duration (Alais et al., 2010).

1.1. An exceptional pattern of audio-visual integration

One exception to this otherwise consistent framework can be found in a musical illusion in which professional marimbists use visible gestures to control the perceived duration of musical notes (Schutz & Lipscomb, 2007). Because this effect is not the result of a response bias or cognitive correction (Schutz & Kubovy, 2009), it represents a curious exception to an otherwise lawful pattern of optimal integration. Subsequent research has demonstrated that this exception is not related to uni-modal information quality, but rather the cross-modal causal link between the striking gesture and the resultant impact sound. Manipulations breaking this link demonstrate no visual influence (Armontrout et al., 2009; Schutz & Kubovy, 2009), consistent with the theory of optimal integration.

1.2. Autism and audio-visual integration

The importance of sensory integration in our everyday experience becomes clear when considering special populations with sensory integration dysfunction. For instance, it is well documented that individuals with autism spectrum disorders (ASD) have difficulties with auditory-visual integration. For example, they struggle with using lip movements to improve speech perception (Smith & Bennetto, 2007), matching faces and voices (Loveland et al., 1995), and integrating communicative gestures with co-expressive speech (Silverman, Bennetto, Campana, & Tanenehaus, 2010). Despite these difficulties, the full extent of sensory integration dysfunction in ASD is unclear. For example, research examining low-level integration of auditory beeps with visual flashes suggests intact audio-visual integration abilities – at least in the context of simple non-social stimuli (Keane et al., 2010; van der Smagt et al., 2007). However, such work generally involves artificial synthesized sounds and stationary visual images, which are not necessarily indicative of real world perceiving (Gaver, 1993). Therefore, we would like to use the Schutz-Lipscomb task to explore audio-visual integration in ASD, as it allows for rigorous analysis of the effect of naturalist visual stimuli (i.e. biological motion) on the perception of ecologically common auditory events (i.e. impact sounds). Furthermore, given that it lacks social and communicative demands, it is well suited for use in ASD populations.

To date, all work in this illusion has used the same procedure and instructions, varying only the stimuli. Participants have always been asked to make judgments of (1) the duration of the heard sound independent of the seen gesture, and (2) the degree to which the duration of the seen gesture "agrees with" the duration of the heard sound.

In order to reliably use this task when working with ASD individuals, we must ask participants to rate the duration of both the musical note and the accompanying striking gesture separately. This differs from previous approaches, which asked participants to rate the duration of the note and the level of *agreement* between the gesture and the note. This modification is an important precursor to testing ASD individuals who often have trouble shifting cognitive set. Therefore when working with this population it is preferable to use a task involving consistent judgments of one dimension (duration), rather than shifting between multiple dimensions (i.e. duration and agreement).

This new procedure differs from previous work by using a task that explicitly calls attention to the gesture's length when judging tone duration. Although the illusion is robust in the face of manipulations preserving the causal link between auditory and visual stimuli, we felt it necessary to first determine whether this alteration in instructions changes the magnitude of the illusion. This is an important consideration, given that attention can diminish task-irrelevant information's influence (Schwarz & Clore, 1983).

1.3. Drawing attention to task-irrelevant information

In order to explore the effect of attentional focus in this paradigm, we conducted two experiments using (1) the original instructions asking about the degree of agreement between the observed gesture and the heard sound, (2) the new instructions calling attention to the length of the observed gesture. This serves as an important precursor to our subsequent investigations of audio-visual integration amongst children with autism. It also explores whether the illusion is robust to manipulations of attention, which will provide further insight into an illusion at odds with the otherwise widely-accepted theory of optimal integration.

2. METHODS

We recruited 42 undergraduate students and ran them in one of two experiments. Two participants who did not appear to be attentive to the task instructions were dropped from subsequent analysis, leaving 40 participants. Although both experiments asked participants to rate the duration of the heard tone, they differed in their instructions regarding a second question about the observed gesture. The first experiment used the original wording ("indicate how much the length of sound 'agrees' with the length of the gesture"), whereas the second experiment used new wording ("indicate how long each gesture looks"). In all other respects the two experiments were identical. The auditory component of the stimuli consisted of the sound produced by either the "long" striking gesture or the "damped" sound created by muffling the bar shortly after striking. The visual component consisted of long and short striking motions produced by a world-renowned marimbist (Schutz & Lipscomb, 2007).

3. RESULTS

Duration ratings were assessed with a 2 (gesture) x 6 (marimba tone) x 2 (question style) repeated-measures ANOVA with *gesture* and *marimba tone* as within-participants variables and *question style* as a between-participants variable. The most important finding (shown in Fig 1) was that question style did not have a significant influence on ratings of tone duration ($F_{1,38}$ =.6664, p = 0.4194), nor was there a question-style x gesture interaction ($F_{1,38}$ =.6401, p =.4286). As expected, there was a main effect of gesture ($F_{1,38}$ = 42.333, $p < 0.0001$).

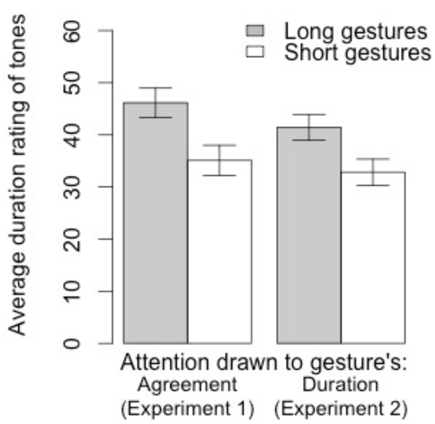

Figure 1. Duration ratings as a function of attentional focus. Vision's influence on perceived tone duration did not differ between presentation conditions (i.e. experiments). Error bars represent 95% confidence intervals about the mean.

4. DISCUSSION AND CONCLUSIONS

Previous research has shown that explicitly drawing attention to task-irrelevant information can reduce its influence (Schwarz & Clore, 1983). However, it appears that explicitly calling attention to the duration of the observed gesture does not compromise the magnitude of this particular illusion. This demonstrates that the illusion is robust with respect to the manipulation of instructions. Although not entirely unexpected, this outcome nonetheless provides an important step forward in our broader goal of adopting this paradigm for using in clinical settings – such as exploring sensory integration dysfunction in children with autism. Given the clear conflict between this illusion and current theoretical viewpoints (predicting minimal visual influence on a temporal duration task), the robust nature of this illusion is also informative with respect to theories of audio-visual integration.

Illuminating the extent of sensory integration abilities and impairments in ASD has the potential for helping us to better understand the underlying mechanisms that may contribute to social and communicative difficulties at the heart of the disorder. In addition, studies delineating aspects of high- and low-level integration and the processing of biological motion could help to clarify the neural basis of the disorder by providing the groundwork for future imaging studies.

In conclusion, the "marimba illusion" provides a rare opportunity to test audio-visual integration that involves biological motion but lacks social and communicative demands. Therefore, this adapted methodology demonstrates an important first step towards using this paradigm as a new tool for novel explorations that assess sensory integration disorder in ASD. Furthermore, the modified procedure described in this abstract makes the task more accessible to a wider range of individuals with varied cognitive skills and social/communicative abilities, opening up many possibilities for future use in studying sensory integration dysfunction amongst a wide range of clinical populations.

REFERENCES

Alais, D. et al. (2010). *Seeing and Perceiving*.
Armontrout, J., et al. (2009). *Attention, Perception, & Psychophysics*.
Gaver, W. (1993). *Ecological Psychology*.
Keane, B. et al. (2010). *Research in Autism Spectrum Disorders*.
Loveland, K. et al. (1995). *Development and Psychopathology*.
Schutz, M. & Kubovy, M. (2009). *JEP:HPP*.
Schutz, M. & Lipscomb, S. (2007). *Perception*.
Schwarz, N., & Clore, G. L. (1983) *J Personality and Social Psychology*.
Silverman, L et al. (2010). *Cognition*.
Smith, E. G., & Bennetto, L. (2007). *J Child Psychology and Psychiatry*.
van der Smagt, M. (2007). *J Autism and Developmental Disorders*.

ACKNOWLEDGEMENTS

Support for this project was provided in part by the NSERC (RGPIN/386603-2010), the Ministry of Research and Innovation's Early Research Award (ER10-07-195), and the NIH (1R03 PAR-07-287). The authors would also like to acknowledge Allison Canfield's helpful contributions to this study.